A note on Severity and Significativity

Salim Lardjane

A note on Severity and Significativity

Salim Lardjane [1]

[1] *LMBA UMR CNRS 6205, Université de Bretagne Sud*
Centre Yves Coppens, Bat. B, 1er ét., Campus de Tohannic
BP 573, 56017 Vannes, France
salim.lardjane@univ-ubs.fr

Abstract. The author compares, in the context of a class of significance tests, Deborah Mayo's notion of evidence, which is based on severity (Mayo 1996, 2004, 2010, 2011) to Bill Thompson's bivariate notion of evidence (Thompson 2007), which is based on significativity. He concludes in favor of the severity approach.

Keywords. Severity, Significativity, Significance tests.

1 Introduction

This work deals with the test of a simple hypothesis against a simple hypothesis, that is, of

$$H_0 : \theta = \theta_0$$

against

$$H_1 : \theta = \theta_1$$

where θ is the true value of the parameter of the common distribution of an i.i.d. sample of observations X. We assume that this distribution admits a density f_θ with respect to a reference measure λ (usually the counting or Lebesgue measure).

The likelihood ratio

$$r(X) = f_1(X)/f_0(X)$$

is retained as a test statistic, where f_0 and f_1 the density of X under assumtion H_0 and H_1 respectively.

Let us denote by x a realization of X.

The degree of significativity or *p-value* of the test for H_0 is given by

$$p(H_0, H_1, x) = \int_{r(u)>r(x)} f_0(u)du.$$

The degree of significativity of the test for H_1 is given by

$$p(H_1, H_0, x) = \int_{r(u) \le r(x)} f_1(u) du.$$

The decision rule consists in accepting H_1 if $p(H_0, H_1, x) \le \alpha$ and accepting H_0 otherwise, where α is a fixed positive scalar, generally "small" (typically, $\alpha = 0.05$).

2 Severity

One says that the hypothesis H_0 passes a test of severity s on data x (Mayo 1996, 2004, 2010, 2011) if (i) x leads to accept H_0 and (ii) the probability of observing a sample which fits at most as well as x to H_0 is equal to s under H_1.

x' is is a worse fit to H_0 than x if $r(x') > r(x)$ and a worse fit to H_1 than x if $r(x') < r(x)$.

For all the samples x *supporting* H_1, one sets

$$s(H_0, H_1, x) = \int_{r(u) \le r(x)} f_0(u) du.$$

For the samples x *supporting* H_0, one sets

$$s(H_0, H_1, x) = \int_{r(u) > r(x)} f_1(u) du.$$

The function $s(H_0, H_1, x)$ thus defined is called the *severity of the test on the data* x.

We shall say that the sample x is a decisive evidence (for the hypothesis it supports) if the severity of the test on x is high (higher than a threshold $1 - \alpha$).

Note that if x supports H_1,

$$s(H_0, H_1, x) = 1 - p(H_0, H_1, x) > 1 - \alpha$$

for our test, and that, if x supports H_0,

$$s(H_0, H_1, x) = 1 - p(H_1, H_0, x).$$

It is thus seen that the severity of a significance test with level α *that leads to a rejection of the null hypothesis* is necessarily higher than $1 - \alpha$, as already emphasized by Mayo (Mayo 1996, p. 194).

A significance test which leads to accept the alternative hypothesis is thus *automatically* a severe test on the observed data, provided the significativity level is low.

3 Bivariate evidence

Bill Thompson (Thompson 2007) quantifies evidence for significance tests by the pair of values

$$ev(x) = (p(H_0, H_1, x), p(H_1, H_0, x)).$$

In the framework of this approach, x is a decisive evidence for H_1 if $p(H_0, H_1, x)$ is low ($< \alpha$) and $p(H_1, H_0, x)$ is high ($> 1 - \alpha$).

Bill Thompson's approach emphasizes the need to *complement* the result of the significance test by an evidence assessment even when it ends up in rejecting the null hypothesis.

This *contradicts* the approach based on severity, in which a rejection of the null hypothesis *automatically* constitutes a decisive evidence in favor of the alternative hypothesis.

4 Severity and Bivariate Evidence

The quantities computed in the case of a rejection of the null hypothesis ($p(H_0, H_1, x) < \alpha$) are the following:

- In the severity-based approach : The probability of obtaining a result as favourable or less favourable to the alternative hypothesis *under the null hypothesis*,

$$s(H_0, H_1, x) \quad = \quad \int_{r(u) \leq r(x)} f_0(u)du = 1 - p(H_0, H_1, x)$$

- In the significativity-based approach : The probability of obtaining a result as favourable or less favourable to the alternative hypothesis *under the alternative hypothesis*,

$$p(H_1, H_0, x) = \int_{r(u) \leq r(x)} f_1(u)du$$

When *severity* is high, the computed likelihood ratio is atypical under the null hypothesis, thus one has very few chances to be wrong when rejecting the null hypothesis.

When the *significativity* of the test for the alternative hypothesis is high, then the likelihood ratio is also atypical for the alternative hypothesis.

Thus, retaining Thompson's criterion as a measure of evidence leads to retain as decisive evidence for the alternative hypothesis likelihood ratios which are atypical under the null hypothesis *and* the alternative hypothesis, which seems unduly restrictive and actually *casts doubt on both hypotheses.*

Let us note that in the example provided by (Thompson 2007, p. 108), Bill Thompson rejects the null hypothesis with an evidence of $(0.05, 0.76)$; but, if a significativity of 0.05 is considered to be low, than a high significativity should correspond to a value of 0.95 at least; there is a contradiction, unless one adopts independent threshold values for low and high significativities. The approach based on severity leads to the conclusion given by Bill Thompson without requiring such an independence.

5 Conclusion

To have a decisive evidence in favour of the alternative hypothesis, one needs, in the severity-based approach, to observe a likelihood ratio which is atypical under the null hypothesis. In the bivariate evidence framework, one needs to aberve a likelihood ratio which is atypical under the null hypothesis *and* the alternative hypothesis. This appears to be unduly restrictive and corresponds to situations where none of the hypotheses is supported by the data. The approach based on severity appears to correspond more closely to inference as it is practised daily in the field of scientific research.

References

[1] Mayo D. G. (1996). *Error and the growth of experimental knowledge*, The University of Chicago Press.

[2] Mayo D. G. (2004). *An error-statistical philosophy of evidence* in *The nature of scientific evidence* (Tapper M. L. and Lele S. R., eds), The University of Chicago Press.

[3] Mayo D. G., Spanos A. (2010). *Error and Inference*, Cambridge University Press.

[4] Mayo D. G., Spanos A. (2011). *Error Statistics* in *Philosophy of Statistics* (Bandyopadhyay P. S. and Forster M. R., eds), Elsevier.

[5] Thompson B. (2007). *The nature of statistical evidence*, Springer.